Reviewer 2

Note: Reviewers' comments are shown in blue, **our responses** in bold black, and *revised manuscript text quoted in the responses* in black italics.

Referee comment

The paper describes a coupled model approach to study marine-terminating glaciers. The workflow joins three established models for climatic mass balance and ice dynamics to estimate mass balances and frontal ablation. The model framework makes use of four tunable parameters, which are tuned in a thorough Bayesian calibration. The manuscript is well structured and written. The individual model components are only briefly described in the manuscript, as they are well documented in the original model papers. However, there are a few important assumptions that should be stated within the manuscript (see specific comments). The overall presentation and figure quality are good, with a few minor adjustments needed.

Response: We thank the reviewer for the encouraging feedback and helpful suggestions. We have revised the manuscript accordingly, for example by clarifying the terminology, updating the figures and captions, adding a workflow diagram, and revising the discussion of surge-type glaciers. Detailed responses to each comment are provided below.

General remarks

Terminology

The title refers to surface mass balance, whereas the manuscript predominantly uses climatic mass balance for the PyGEM component. I assume that these are used interchangeably here, since the PyGEM formulation described includes accumulation, surface melt/ablation, and refreezing, while internal and basal mass- balance components are neglected. If this is the intended meaning, I suggest stating this explicitly when climatic mass balance is first introduced, to avoid ambiguity.

Response: Thank you for spotting it. PyGEM computes the climatic mass balance including refreezing; therefore the term climatic mass balance is used in the entire manuscript. The title should also be climatic mass balance and we revised it accordingly.

Ensemble: Make sure that it is clear throughout when "ensemble" is mentioned whether the context refers to the Bayesian ensemble or the ensemble of climate models (for example, Fig. 1). Especially in the figures, I would rather have it stated one more time than not. In the manuscript text it is usually clear.

Response: Thank you for this suggestion. We have revised the relevant figure and table captions to explicitly distinguish the Bayesian ensemble from the ensemble of 12 global climate models (GCMs):

- Figures 2, 3, S1, and S6: We added, *"The gray lines represent individual members of the Bayesian ensemble."*
- Figures 4 and S2 and Table S2: We now explicitly use the term *"Bayesian ensemble."*
- Figures 5, S3, and S4: We added, *"Each boxplot summarizes the projected total length change across the ensemble of 12 GCMs for a given glacier and scenario."*

- Figure 6: We revised the caption to state, *"Solid lines and shaded areas depict the median and interquartile range (25th–75th percentiles, IQR) across the ensemble of 12 GCMs for each scenario, respectively."*

Specific remarks

Line 77ff: Detailed information can be found in the referenced papers, but it is not clear a) if the temperature bias correction is acting on the threshold for solid/liquid precipitation, refreezing, and melting, or only on some of these processes. Based on the references, if this is true, change it to the "climate/ERA5" model temperature bias. This makes it clear where in the model chain the bias correction is applied. b) State that the degree-day factors of ice and firn are dependent on the one for snow, and therefore only the snow degree-day factor is calibrated. In the current manuscript, the reader could easily believe that the other two are kept constant. (Section 3.1 in Rounce et al. 2020a)

Response: As suggested, we have added the following clarification to address both points (**Lines 78-80**):

"The degree-day factors for ice and firn are linked to that for snow by a fixed ratio and therefore vary with f_{snow} rather than being calibrated independently. The additive correction T_{bias} is applied to the climate temperature forcing (ERA5) used to calculate precipitation partitioning, melting, and refreezing; the rain-snow temperature threshold remains fixed."

Lines 125ff: A small flowchart would increase understandability and serve as a quick reference during the reading process.

Response: we have added a workflow diagram illustrating the coupling between PyGEM, OGGM, and SERMeQ (Figure 1; reproduced below), and cited it in the main text (**Line 130**) to provide a clear visual reference: *(Figure 1)*.

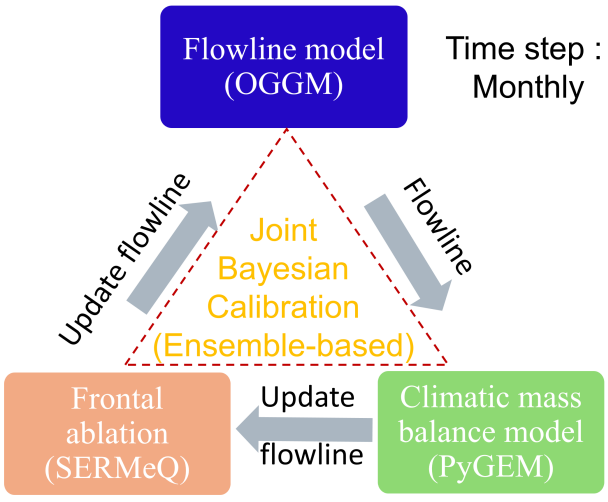


Figure 1. Workflow of the coupled PyGEM–OGGM–SERMeQ model. SERMeQ is called within the OGGM flowline model.

Section 2.3: Moving the extended description of your Bayesian calibration method to the supplement is good. Just make sure that all main assumptions of your method are given in the

main manuscript. "For simplicity we have here made the typical strong constraint assumption (see Evensen et al., 2022) that the forward model is perfect other than errors in the parameters." This may need to be stated in the main text prior to/around line 144.

Response: As suggested, we have moved the statement on the strong-constraint assumption to Section 2.3 of the main manuscript (**Lines 146-147**). We have also revised the supplementary text to refer back to this statement, while retaining the supporting discussion and references.

Line 250: State which three variables are meant; on first read, it was not immediately clear to me, and I had to go to the figure and back to the lines above.

Response: As suggested, we now explicitly identify the three variables in the revised text (**Lines 259-260**): *"total length change over the period, period-averaged climatic mass balance, and period-averaged frontal ablation."*

Figure 3 and corresponding text: There are outliers with much worse metrics. Maybe the plot could be altered to include symbols/color shading for each glacier, so one could see if it is the same ones that deviate strongly from the 1:1 line in 3a, 3c, and 3e. If so, only highlighting those could also be an option. (Do we know why?) The manuscript text need not contain a full discussion of those, but it would be possible to discuss whether the "bad" performance occurs for each variable or only one.

Response: We compared the largest mismatches across length change, climatic mass balance, and frontal ablation for the 62 glaciers with converged calibration. Mismatches were ranked by the separation between the modeled ensemble range and the observation interval defined by the observed value plus or minus its reported uncertainty. The glaciers with the largest mismatches are partly shared across the three variables, but their rankings differ among variables. We also explored glacier geometry and RGIv6 surge classification. No consistent geometry-related pattern emerged across variables and periods; surge-type glaciers were more frequently represented near the top of the mismatch rankings in several comparisons, although this varied among variables and periods.

To preserve the readability of the already dense Figure 3, we have retained its existing format and added the following clarification in Section 4.2 (**Lines 268-276**), and discuss potential contributors and limitations in Sections 6.1 and 6.3:

"Among glaciers with converged calibration, mismatches were ranked separately for each variable and period by the separation between the modeled ensemble range and the observation interval, defined as the observed value plus or minus its reported uncertainty. The glaciers with the largest mismatches were partly shared across length change, climatic mass balance, and frontal ablation, but their rank order differed among variables. An exploratory comparison with glacier geometry attributes from RGIv6 (area, maximum length, mean surface slope, and median elevation) did not reveal a consistent geometry-related pattern across all three variables and both the calibration and validation periods. Additionally, glaciers classified as surge-type in RGIv6 (RGI Consortium, 2017) were more frequently represented near the top of the mismatch rankings in several comparisons, although the patterns varied among variables and periods. Potential contributors to the mismatches are discussed in Section 6.1, and the implications of surge classification are considered in Section 6.3."

Additionally, we also add the clarification about limitations of calibration data and revised the 6.1 section as (**Lines 335-342**):

"Despite these strengths, limitations remain at the glacier scale. Frontal ablation is systematically overestimated, highlighting shortcomings in how terminus dynamics are represented and constrained. This bias may arise from compounded uncertainties in the initial flowline geometry (e.g., bed, thickness, and width), errors in derived velocities, and the fact that velocity is only

indirectly constrained by observations. Frontal ablation observations are also not directly used as calibration targets. In addition, the annual length change records for some glaciers are extensively infilled by interpolation, limiting the amount of directly observed information available to constrain year-to-year variability. Together, these limitations point to the need for more accurate bed inversions, better constrained velocity estimates, and more complete observational coverage of annual length change to improve glacier-scale model performance."

Section 5, Lines 211ff/Section 6.2: Please clarify whether the parameters obtained from the Bayesian calibration using ERA5 are also used for the simulations driven by the individual climate models, or whether the parameters are recalibrated separately for each climate model. In particular, I assume that the calibrated temperature bias correction (Tbias) and precipitation correction factor (kp) obtained with ERA5 are applied unchanged to all climate models. If this is the case, this assumption should be stated explicitly.

Using the same bias-correction parameters for all climate models and ERA5 is a simplification, since the magnitude and structure of model-specific biases may differ between the climate models. However, given that the primary purpose of the paper is to demonstrate the Bayesian calibration methodology and that the projections are presented mainly as a proof of concept, I consider this assumption reasonable. Nevertheless, it should be clearly stated and discussed as a limitation.

Response: Yes, all four parameters calibrated using ERA5 are retained without recalibration for individual GCMs. The GCM forcing is separately bias-adjusted to ERA5. We have clarified these two steps in Section 5 (**Lines 278-281**):

"All four parameters obtained from the Bayesian calibration using ERA5 are applied unchanged in projections driven by all 12 GCMs under both scenarios, without recalibration for individual GCMs. Each GCM's temperature and precipitation forcing is separately bias-adjusted using ERA5 as the reference, following the PyGEM procedure, before applying the calibrated model parameters."

As suggested, we have also acknowledged the associated assumption and limitation in Section 6.2 (**Lines 351-353**):

"Applying the parameters calibrated using ERA5 unchanged to all simulations driven by GCMs assumes that they remain applicable after bias adjustment to ERA5, although residual biases specific to individual GCMs may still affect the projections."

Figure 4: If this figure contains glaciers which are excluded later in the process (L 312 and L 322) for the analysis, highlight those in the figure. It could well be that they are not calculated at all and are therefore not shown here anyway, but I cannot fully figure this out. Does the cumulative mass change (Fig. 5) include the unconverged ones? Clarify which glaciers are excluded where and why more clearly.

Response: In Figures 5, 6, S3, S4, we did not include the four glaciers discussed in Section 6.2 (L 312 and L 322), which were excluded before calculating the projection statistics. The 67 retained glaciers comprise 59 with converged calibration and 8 with unconverged calibration. All 67 retained glaciers are included in the regional totals in Figure 6. We have added the following clarification at the end of Section 5 (**Lines 302-305**), with references to Table S1 for the glacier identities and Section 6.2 for the detailed exclusion reasons:

"The projection results presented in this section and in Figures 5, 6, S3, and S4 are based on 67 of the total 71 glaciers. Four glaciers (Table S1) were excluded before calculating the projection statistics because of the physical instabilities and model limitations discussed in Section 6.2. The retained sample comprises 59 glaciers with converged calibration and 8 with unconverged calibration. All 67 retained glaciers contribute to the regional totals shown in Figure 6."

Why are you plotting only the length change per glacier (Fig. 4), while climatic mass balance and frontal ablation are only shown cumulatively (Fig. 5)?

Response: As suggested by both reviewers, we have added figures showing cumulative climatic mass balance and frontal ablation over 2000–2100 for each of the 67 glaciers in Svalbard as **Figure S3 and Figure S4**, respectively, in the **Supplementary** file (reproduced below). These figures follow the map and boxplot format of Figure 4. We have also added the following description and figure references to the **main** text (**Lines 289-292**):

"For the same glaciers, the median cumulative climatic mass balance over 2000-2100 is generally lower under SSP585 than under SSP126, refelcting smaller gains where the climatic mass balance is positive and greater losses where it is negative (Figure S3). Scenario differences in median cumulative frontal ablation are less uniform, although most glaciers show slightly higher values under SSP585 (Figure S4)."

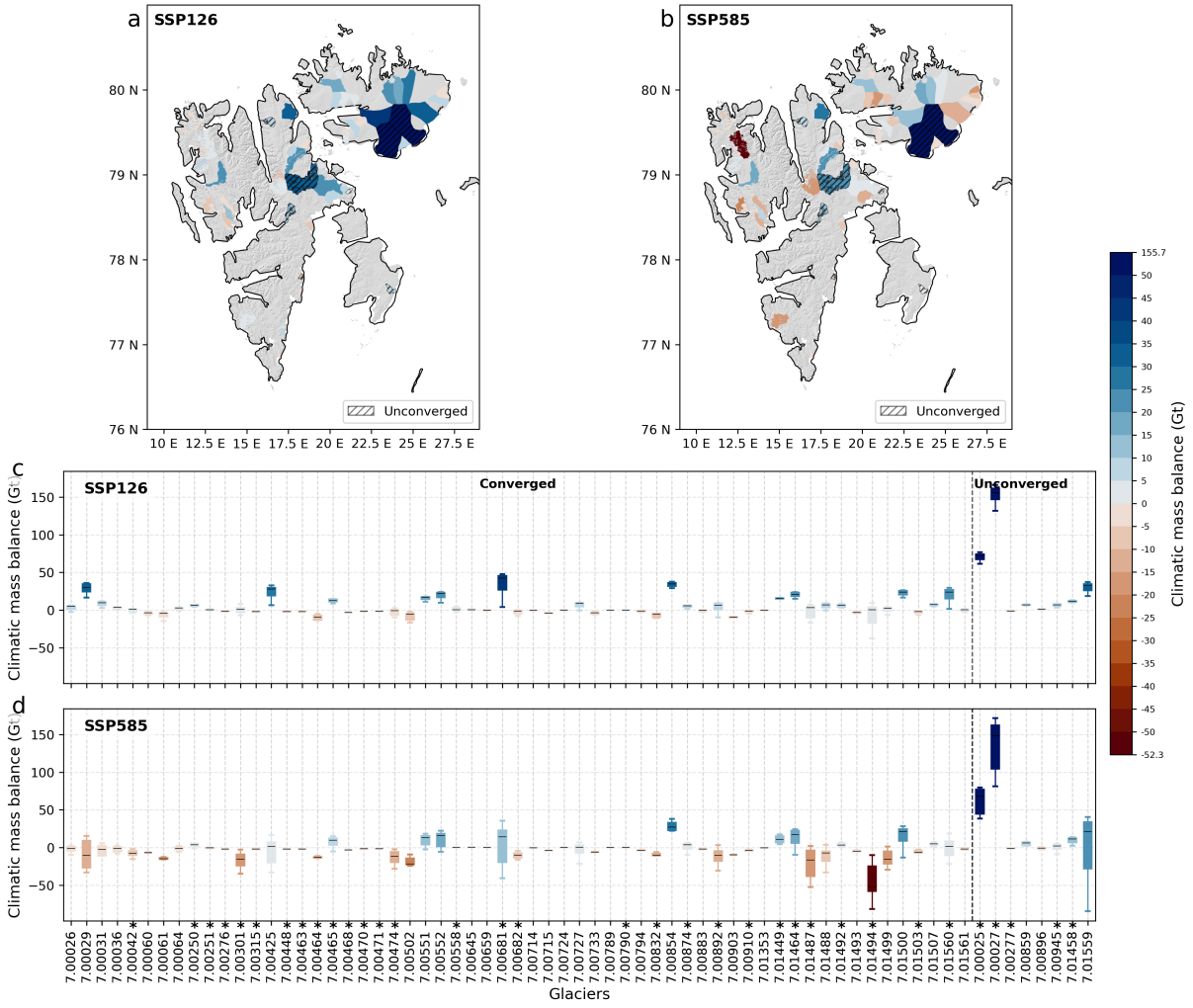


Figure S3. Projected cumulative climatic mass balance over 2000-2100 for each of the 67 glaciers in Svalbard. (**a-b**) Spatial distribution of median cumulative climatic mass balance under SSP126 and SSP585 scenarios, with hashed areas indicating unconverged calibrations (9 glaciers). (**c, d**) Distribution of projected cumulative climatic mass balance of each glacier under scenarios SSP126 and SSP585, grouped by glaciers with converged calibrations (left) and unconverged (right). Each boxplot summarizes the projected total length change across the ensemble of 12 GCMs for a given glacier and scenario. Box elements: the center line indicates the median, the box spans the 25th–75th percentiles (interquartile ranges, IQR), and the whiskers extend to 1.5×IQR. Color scale represents magnitude of climatic mass balance, facilitating comparison between spatial and statistical representations. Glaciers are identified by their RGIv6 IDs listed without the common "RGI60-0" prefix, and surge-type glaciers are marked with an asterisk (*).

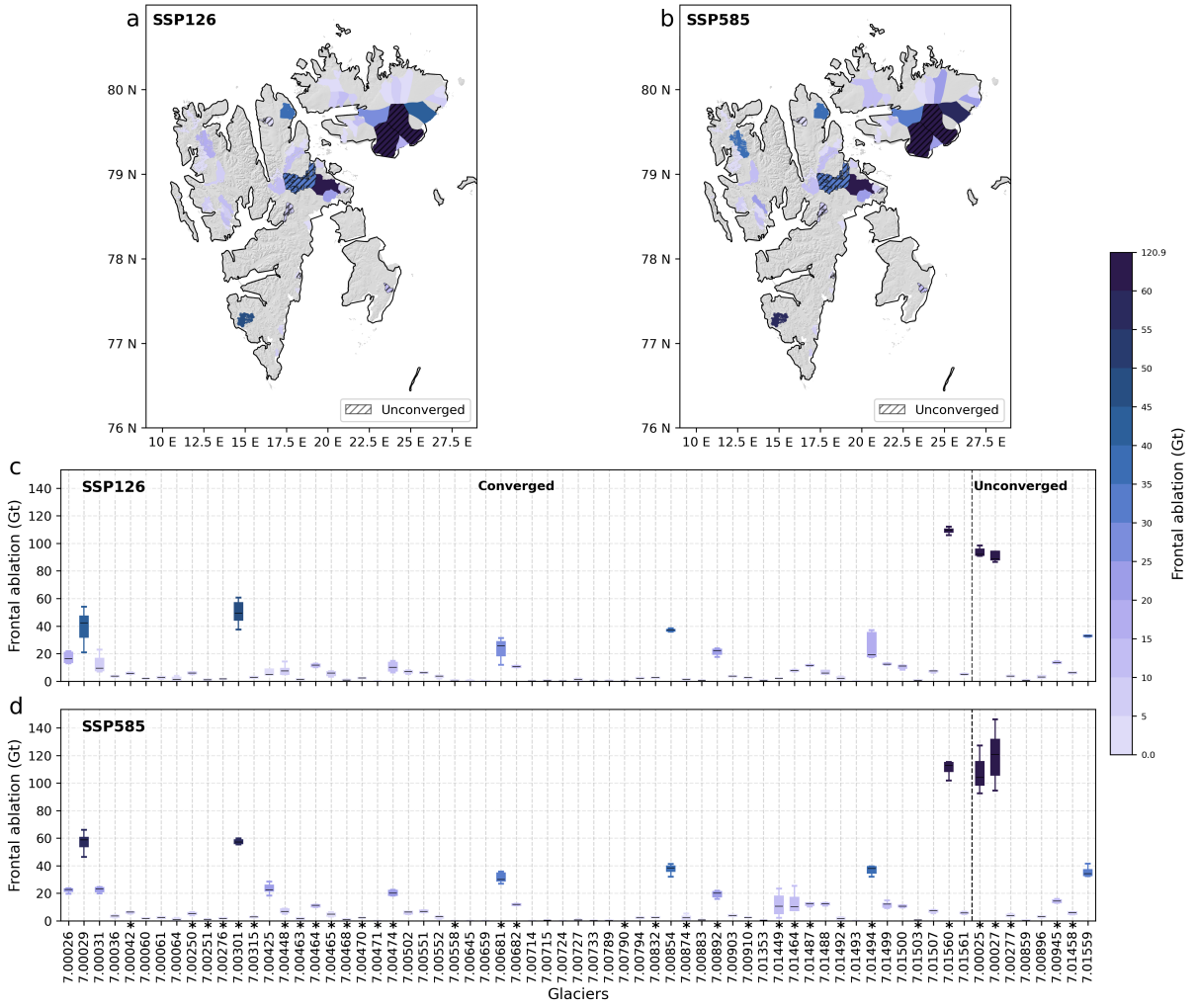


Figure S4. Projected cumulative frontal ablation over 2000-2100 for each of the 67 glaciers in Svalbard. (**a-b**) Spatial distribution of median cumulative frontal ablation under SSP126 and SSP585 scenarios, with hashed areas indicating unconverged calibrations (9 glaciers). (**c, d**) Distribution of projected cumulative frontal ablation of each glacier under scenarios SSP126 and SSP585, grouped by glaciers with converged calibrations (left) and unconverged (right). Each boxplot summarizes the projected total length change across the ensemble of 12 GCMs for a given glacier and scenario. Box elements: the center line indicates the median, the box spans the 25th–75th percentiles (interquartile ranges, IQR), and the whiskers extend to 1.5×IQR. Color scale represents magnitude of length change, facilitating comparison between spatial and statistical representations. Glaciers are identified by their RGIv6 IDs listed without the common "RGI60-0" prefix, and surge-type glaciers are marked with an asterisk (*).

Section 6.3: Having a substantial number of surge-type glaciers in the dataset makes the entire approach less robust in its findings. Can you comment and add to what degree surge processes are occurring during the "training period" beyond the statement that they "showed limited influence"? Add a discussion about how, if any surges occur during the 19 years, the terminus movement actually impacts the calibration approach.

Response: We agree that our original statement "showed limited influence" was insufficiently supported. We have removed it and revised Section 6.3. In this application, 2000–2010 is used for calibration and 2010–2020 for validation. The RGIv6 surge classifications used in this study do not provide event timing or magnitude. Our analysis does not establish the extent of active

surging during these periods. This model-description study focuses on integrating the SERMeQ frontal-ablation formulation into the PyGEM–OGGM framework with joint Bayesian calibration and demonstrating its regional application using Svalbard as a test region. Incorporating explicit surge dynamics and reproducing surge events or cycles lie beyond the scope of this study.

However, differences in model performance between glaciers classified as surge-type in RGIv6 and those with no surge evidence warrant examination. Surge-related terminus movements may influence calibration through the length-change observations, potentially causing parameter adjustments to compensate for processes not explicitly represented by the framework. Whether the observed differences arise from active surge processes, and whether and how explicitly representing those processes would improve model performance, remain open questions.

We have identified surge-type glaciers in Table S1, reported convergence proportions for both groups in Section 6.3, and compared observation–model mismatches by surge classification in Section 4.2. The revised Section 6.3 discusses these limitations and acknowledges that the effects of unrepresented surge processes on regional projections remain unquantified.

The following text has been added to Section 4.2 (**Lines 273-276**):

"Additionally, glaciers classified as surge-type in RGIv6 (RGI Consortium, 2017) were more frequently represented near the top of the mismatch rankings in several comparisons, although the patterns varied among variables and periods. Potential contributors to the mismatch are discussed in Section 6.1, and the implications of surge classification are considered in Section 6.3."

The revised Section 6.3 reads as follows (**Lines 367-381**):

"A high proportion of Svalbard's marine-terminating glaciers are classified as surge-type in RGIv6 (RGI Consortium, 2017). Glacier surges are typically viewed as internal instabilities (Meier and Post, 1969), characterized by cyclical flow between extended quiescent periods and brief surge events. Thirty-six out of 71 glaciers in this study were classified as surge-type in RGIv6 (RGI Consortium, 2017) and are identified in Table S1. Under the prescribed hyperparameter configuration, 30/36 surge classified glaciers (83.3%) and 32/35 glaciers with no surge evidence (91.4%) converged within 10 iterations (Table S1). Among these converged cases, the mismatches reported for surge-type glaciers in Section 4.2 do not establish that active surging caused the discrepancies.

Our frontal-ablation model uses seasonally to annually averaged stress fields without explicit subglacial hydrology, and the coupled framework does not explicitly represent surge events or cycles. It can nevertheless simulate terminus change and frontal ablation for surge-type glaciers. Surge-related terminus movements may influence calibration through the length-change observations, potentially causing parameter adjustments to compensate for processes not explicitly represented by the framework. However, RGIv6 classifications alone do not provide the event timing and magnitude needed to assess these effects. The results therefore illustrate the framework's application under the selected settings, while the need for additional treatment of surge-classified glaciers remains open. Future projections should be interpreted with the caveat that unrepresented surge cycles may alter terminus positions, and their consequences for regional projections have not been quantified."

References

Meier, M. F. and Post, A.: What are glacier surges?, Canadian Journal of Earth Sciences, https://doi.org/10.1139/e69-081, publisher: NRC Research Press Ottawa, Canada, 1969.

RGI Consortium: Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 6.0, Technical Report, Global Land Ice Measurements from Space, 2017.